Investigation of impact of shoreline alteration on coastal hydrodynamics using Dimension REduced Surrogate based Sensitivity Analysis

Jia, Gaofeng; Wang, Ruo-Qian; Stacey, Mark T.

Published in:
Advances in Water Resources

DOI:
10.1016/j.advwatres.2019.03.001

Publication date:
2019

Investigation of Impact of Shoreline Alteration on Coastal Hydrodynamics using Dimension Reduced Surrogate based Sensitivity Analysis

Article *in* Advances in Water Resources · March 2019

DOI: 10.1016/j.advwatres.2019.03.001

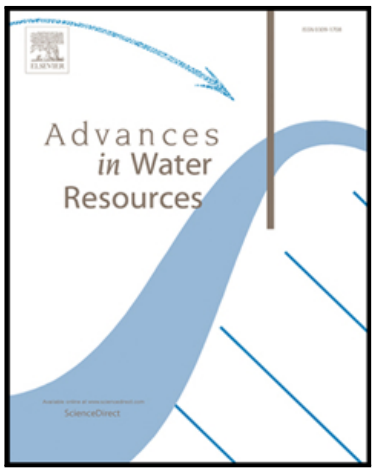
CITATIONS
0

READS
73

3 authors, including:

Ruo-Qian Wang
Rutgers, The State University of New Jersey
48 PUBLICATIONS **138** CITATIONS

SEE PROFILE



Investigation of Impact of Shoreline Alteration on Coastal Hydrodynamics using Dimension Reduced Surrogate based Sensitivity Analysis

Gaofeng Jia, Ruo-Qian Wang, Mark T Stacey

PII: S0309-1708(18)30526-8
DOI: https://doi.org/10.1016/j.advwatres.2019.03.001
Reference: ADWR 3289

To appear in: *Advances in Water Resources*

Received date: 16 June 2018
Revised date: 25 February 2019
Accepted date: 2 March 2019

Please cite this article as: Gaofeng Jia, Ruo-Qian Wang, Mark T Stacey, Investigation of Impact of Shoreline Alteration on Coastal Hydrodynamics using Dimension Reduced Surrogate based Sensitivity Analysis, *Advances in Water Resources* (2019), doi: https://doi.org/10.1016/j.advwatres.2019.03.001

Highlights

- A sensitivity analysis based approach to understand the impact of the variation in the shoreline on the water levels over the coastal region under Sea Level Rise
- Propose Dimension REduced Surrogate based Sensitivity Analysis (DRESSA) method to facilitate efficient sensitivity analysis for expensive high-fidelity numerical models with high-dimensional outputs.
- Principal Component Analysis (PCA) to exploit the correlation in the high-dimensional outputs to find a low-dimensional latent output representation
- Surrogate models for the latent outputs based on a small number of runs of the high-fidelity numerical model
- DRESSA establishes sensitivity indexes and directly transforms for the high-dimensional outputs
- DRESSA is applied to generate sensitivity maps and investigate the impact of different containment strategies on peak water level (PWL) over the entire San Francisco Bay under SLR

Investigation of Impact of Shoreline Alteration on Coastal Hydrodynamics using Dimension Reduced Surrogate based Sensitivity Analysis

Gaofeng Jia[a], Ruo-Qian Wang[b,c], Mark T Stacey[d]

[a]*Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO, USA*
[b]*School of Science and Engineering, University of Dundee, Scotland, UK*
[c]*Department of Civil and Environmental Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ, USA*
[d]*Department of Civil and Environmental Engineering, University of California, Berkeley, Berkeley, CA, USA*

Abstract

To inform the decision-making of coastal protection against sea level rise (SLR), we have to estimate the impact of shoreline alterations on the hydrodynamics. This task involves estimating of a large number of shoreline decision combinations. Here we present a sensitivity analysis based approach to understand how the variation in the shoreline, for example, due to construction of seawalls at different location along the shoreline, would impact the water levels over the coastal region under SLR. To facilitate efficient sensitivity analysis for expensive high-fidelity numerical models with high-dimensional outputs, we propose a Dimension REduced Surrogate based Sensitivity Analysis (DRESSA) method. DRESSA uses Principal Component Analysis (PCA) to exploit the correlation in the high-dimensional outputs to find a low-dimensional latent output representation, then builds a surrogate model for the latent outputs based on a small number of runs of the high-fidelity numerical model. In the end, DRESSA first establishes relevant covariance matrices for the low-dimensional latent outputs using the surrogate model, and then directly establishes sensitivity indexes for the high-dimensional outputs using these covariance matrices and the derived

*Corresponding author
Email address: r.u.wang@dundee.ac.uk (Ruo-Qian Wang)

transformation between sensitivity in latent space and original space. We applied this method to generate sensitivity maps and investigate the impact of different containment strategies on peak water level (PWL) over the entire San Francisco Bay under SLR.

Keywords: Sea-level Rise, Surrogate Assisted Sensitivity Analysis, Flood, Principal Component Analysis (PCA), Data-driven Analysis

1. Introduction

Coastal communities around the world are facing the challenge of mitigating coastal flood risk induced by sea-level rise (SLR). The challenge is particularly acute where retreating from the shoreline is prohibitively expensive or impossible. In these cases, communities frequently act to raise existing levees or seawalls (or construct new ones), which allow them to keep their shorelines, and the infrastructure beyond them, in place. This type of coastal protection infrastructure development can alter local hydrodynamics and spread to the adjacent region [6, 14, 1, 10]. The resulting hydrodynamic interactions transform local actions into regional impacts, particularly in the future under more extreme SLR scenarios [21]. As the impact of SLR on tidal dynamics is receiving increased concerns, there is a need to further our understanding in the hydrodynamic sensitivity to the shoreline alteration to guide the climate adaption activities in the coastal communities. The present paper is aimed at addressing this issue, proposing a systematical method using global sensitivity analysis to examines the impacts of various combinations of interventions involving multiple actors taking protective measures. The information obtained from this study can be used to guide the decision making regarding construction of shoreline infrastructure and also climate change adaptation strategies.

There are multiple challenges in examining the sensitivity of the hydrodynamics to the shoreline alteration. First, high-fidelity numerical models are required to resolve the detailed hydrodynamics around seawalls under SLR. Due

to its high resolution, such numerical models typically take high computational time and costs [20]. Second, a sensitivity analysis typically requires numerous model evaluations considering various parameter combinations. For the problem of investigating multiple actors taking protective measures along large coastal region, the number of combinations could become extremely large [20], e.g., when n_x locations (i.e., actors) are involved in the binary decision of whether or not to take action, it will require a total number of combinations of 2^{n_x} to exhaust all the possibilities. Third, the output of the high-fidelity model has an extremely high dimension, which is difficult to process and analyze. Therefore, direct adoption of the high-fidelity models is not realistic and we need a method to accelerate our examination of all the alternative solutions.

Here we propose surrogate modeling, which uses a database created by running a small number of high-fidelity simulations to train an efficient surrogate model to predict the rest of the combinations that are not in the database. In order to overcome the high-dimensionality difficulty in the simulation outputs, we adopt the dimensionality reduction techniques to reduce the surrogating workload. This efficient dimension reduction and surrogate based approach for sensitivity analysis is named Dimension REduced Surrogate based Sensitivity Analysis (DRESSA). Specifically, DRESSA first uses Principal Component Analysis (PCA) for dimension reduction and establishes a low-dimensional latent output representation to address the challenge of building a surrogate model for high-dimensional outputs, and then builds a surrogate model for latent outputs. DRESSA carries out the sensitivity analysis in two steps: first, it evaluates the relevant covariance matrices for the low-dimensional latent outputs using the established surrogate model, and second, it establishes sensitivity indexes for the high-dimensional outputs using the derived relationship/transformation between these sensitivity indexes and the covariance matrices in the low-dimensional latent output space. We applied DRESSA to investigate how the construction of containment at different locations around San Francisco Bay (SF Bay) would impact the peak water level (PWL) over the

entire SF Bay region under SLR. Sensitivity maps useful for guiding decision making are generated for first order main effects and second order interactions. Though the example focuses on SF Bay, the method is general and could be easily applied to other regions (e.g. Gulf of Mexico [13]) or similar numerical simulation tasks that involves sensitivity analysis of high-dimensional outputs (e.g. the large-scale numerical simulations [19]).

2. Methods

2.1. San Francisco Bay Study Site and Numerical Model

San Francisco Bay (SF Bay) consists of four tidal basins including South SF Bay (SSFB), Central Bay (CB), San Pablo Bay (SPB), and Suisun Bay (SB) (Figure 1). Golden Gate Strait connects SF Bay and the Pacific Ocean. This strait is 100 m deep and has mixed tides dominated by semidiurnal and diurnal constituents. SF Bay is strongly tidal, with tidal amplitudes as high as 1.5 m at the mouth and as much as 60% larger in the interior of the Bay. The major freshwater discharges are the Sacramento and San Joaquin Rivers (Figure 1), which contribute up to 90% of the total freshwater influx [4]. South San Francisco Bay's shoreline has been extensively altered through the construction of levees around most of its perimeter, leading to a narrowing cross section with distance along the Bay's axis. A deep central channel extends the length of SSFB and has a depth of 12-20 m. The rest of the bay is a shallow basin, ranging from 5 m depth to the intertidal plain. Tides in SSFB are close to a standing wave. The velocity phase leads the water level peak by $\sim 75^{\circ}$ [6]. The northern reach of SF Bay is a series of connected tidal basins, starting from Central Bay, to San Pablo Bay and Suisun Bay. The perimeter areas of San Pablo Bay and Suisun Bay were historically extensive tidal marshes, although much of that area is now separated from tidal action by levees or gates. Upstream of these basins lies the Sacramento-San Joaquin Delta, which is a network of channels that serve to effectively dissipate the incoming tidal wave, resulting in

very little reflection along the axis of the northern reach of SF Bay. A deep channel is maintained through these basins and into the delta.

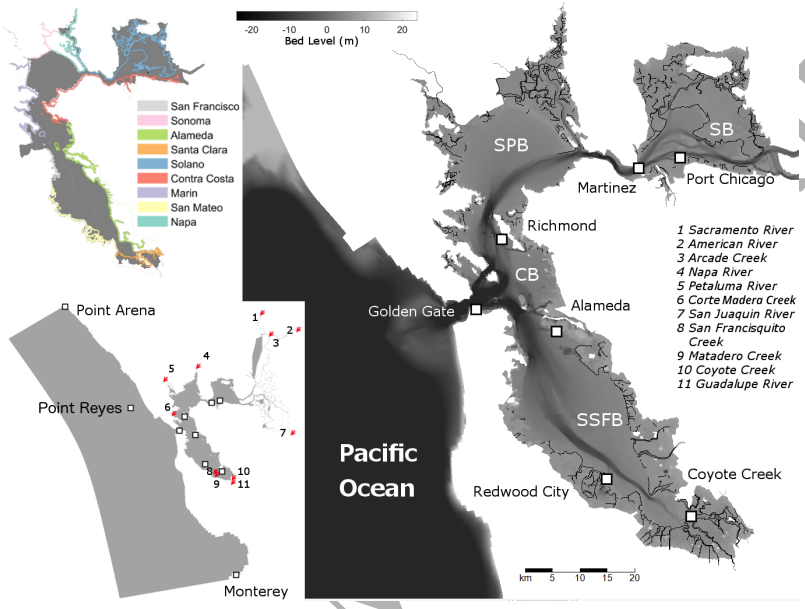


Figure 1: The computational domain covers SF Bay and an open ocean area. The colored lines represent the position of seawalls with different colors for each county (with the legend at top left corner of the figure). The gray dash line represents the position of the existing levees. The labeled number indicates the source location of the freshwater discharge (with name of each number listed on the right side of the figure). The names of the tidal basins in San Francisco (SF) Bay are South SF Bay (SSFB), Central Bay (CB), San Pablo Bay (SPB), and Suisun Bay (SB). The white boxes indicate the observation points in the computational domain that are used to validate the numerical model.

We used Delft3D Flow Flexible Mesh (D-Flow FM) to accurately simulate the tidal dynamics of SF Bay under different sea level rise scenarios [2]. This software solves the shallow-water equation using an unstructured grid. The present paper employed a depth-integrated 2-D grid that was developed and shown reliable through a series of studies [4, 5, 11]. We further increased the resolution up to the scale of 50 m close to the shoreline, allowing us to more accurately resolve the coastal infrastructure and other flood-control features. Up-

dated shorefront topography was added following a recently released database of *Doehring et al.* [3]. Levee structures were simulated using the empirical "fixed weir" model described in the manual of D-Flow FM [2]. This model inserts a flow barrier at the specified location and the geometry of the domain and the momentum transport in the neighboring cells is adjusted using nonlinear empirical models. An open boundary condition was applied outside of the Golden Gate. The north and south sides of the boundary were specified using the tidal constituents at Point Arena and Monterey respectively [12]. The west side was specified by a linear interpolation between these two sites, and was located ∼70 km into Pacific Ocean. The model included 11 river discharges imposed at the numbered locations in Figure 1. The flow rate time series was obtained from the United States Geological Survey gauging sites [18]. A drying and wetting numerical scheme was applied to the intertidal region. The grid was partitioned into 28 parts. MPICH code was used to conduct the parallel computing on a 40 core workstation. A model validation showed that the present numerical model can reproduce the tidal water level with a correlation coefficient higher than 0.98. More details about the model can be found in *Wang et al.* [20].

2.2. Sensitivity Analysis and Sensitivity Maps

To understand how the variation in the shoreline (e.g., construction of containments such as levees, seawalls) at different locations along the shoreline of SF Bay would impact the PWL over the entire bay region under SLR, we use sensitivity analysis. The problem is formulated into the typical form of sensitivity analysis, that is, how the variability/uncertainty in the inputs impacts the variability in the outputs. To facilitate this sensitivity analysis, the formulation treats the decision of whether to build a containment at each considered segment along SF Bay as design variables or inputs to the numerical model, i.e., $\mathbf{x} = [x_1, \ldots, x_i, \ldots, x_{n_x}]^T$ with n_x the considered number of segments. We artificially treat each component x_i in $\mathbf{x}$ as uncertain with uniform distribution in [0, 1]. $x_i < 0.5$ means that the i^{th} containment will not be constructed, and $x_i \geq 0.5$ means that the i^{th} containment will be constructed. For the output

$\mathbf{y} = \mathbf{y}(\mathbf{x})$, we are interested in the PWL at each location in the computational domain, i.e., $\mathbf{y} = [y_1, \ldots, y_k, \ldots, y_{n_y}]$ with n_y the number of locations, which is typically a large number, meaning $\mathbf{y}$ corresponds to extremely high-dimensional output. By looking at how the variability/uncertainty in x_i impacts the variability in the output $\mathbf{y}$ (i.e., PWL), we can establish how the construction of containments at each studied location will impact the water level in the entire bay region. In the end, a sensitivity map for each input x_i can be established.

For sensitivity analysis, Sobol' index, which is the most commonly used variance-based global sensitivity measure, is adopted. According to variance-based sensitivity analysis, the total variance of y due to the variability/uncertainty in $\mathbf{x}$, i.e., $V_y = Var_{\mathbf{x}}[y(\mathbf{x})]$, can be decomposed into the various orders of interaction as $V_y = \sum_{i=1}^{n_x} V_i + \sum_{i=1}^{n_x-1} \sum_{j=i+1}^{n_x} V_{[ij]} + \cdots$, where $V_i = Var_i\left[E_{\sim i}[y(\mathbf{x})|x_i]\right]$ denotes the expected reduction in variance of y (a scalar output) due to fixing x_i and $V_{[ij]} = V_{ij} - V_i - V_j$ where $V_{ij} = Var_{ij}\left[E_{\sim ij}[y(\mathbf{x})|x_i, x_j]\right]$. Here $\sim_i$ denotes the remaining of the model parameter vector $\mathbf{x}$ excluding x_i. Higher order interactions can be expressed similarly. The first order main effect Sobol' index S_i for x_i is defined as [15] $S_i = V_i/V_y$; whereas, the main effect Sobol' index for second order interaction between x_i and x_j is defined as $S_{[ij]} = V_{[ij]}/V_y$. Based on the definition, sensitivity index is unit less and can be understood as a percentage value. For the current problem, we are interested in the sensitivity index for each of high-dimensional outputs. Let $S^y_{k,i}$ denote the first order sensitivity index of the k^{th} output y_k with respect to the i^{th} input x_i, and $S^y_{k,i}$ needs to be calculated for $k = 1, \ldots, n_y$ and $i = 1, \ldots, n_x$. All the sensitivity information can be represented by a $n_y \times n_x$ sensitivity index matrix $\mathbf{S^y}$. Sensitivity index matrix for higher order interactions can be defined similarly. In the end, sensitivity maps can be generated to help visualize the sensitivity index for all outputs.

Calculation of $S^y_{k,i}$ for each i involves evaluation of $Var_i\left[E_{\sim i}[y_k(\mathbf{x})|x_i]\right]$ corresponding to a double integral. The most generalized approach is Monte Carlo

Simulation (MCS) [15]. However, MCS typically entails large number of model evaluations (e.g., N evaluations). Direct adoption of the high-fidelity numerical model in the context MCS is not computationally feasible. In addition, calculation of sensitivity for each of the high-dimensional output (i.e., n_y is large) directly/separately entails high computational effort and memory requirements. For example, if N samples are used for MCS, storing outputs for all samples results to very large output matrices in the order of $N \times n_y \times n_x$.

2.3. Dimension REduced Surrogate based Sensitivity Analysis (DRESSA) for High-dimensional Outputs

To address the above challenges and facilitate efficient sensitivity analysis for expensive models with high-dimensional outputs and generation of sensitivity maps, a Dimension REduced Surrogate based Sensitivity Analysis (DRESSA) method integrated with dimension reduction technique is proposed. The flowchart for the overall method is illustrated in Figure 2, and the following sections provide brief discussions for the steps.

2.3.1. Kriging with PCA for Surrogate Modeling of High-dimensional Outputs

To address the computational challenges associated with sensitivity analysis for expensive high-fidelity numerical models, we use kriging surrogate modeling, while the challenge of building a surrogate model for high-dimensional outputs is addressed through dimension reduction using PCA, which explores the correlation within the outputs and establishes a low-dimensional latent output representation.

The idea of surrogate modeling is establishing a simple mathematical relationship between the inputs and outputs based on a database of high-fidelity numerical results with the ultimate goal to maintain the accuracy of the numerical model utilized to produce this database while providing greatly enhanced com-

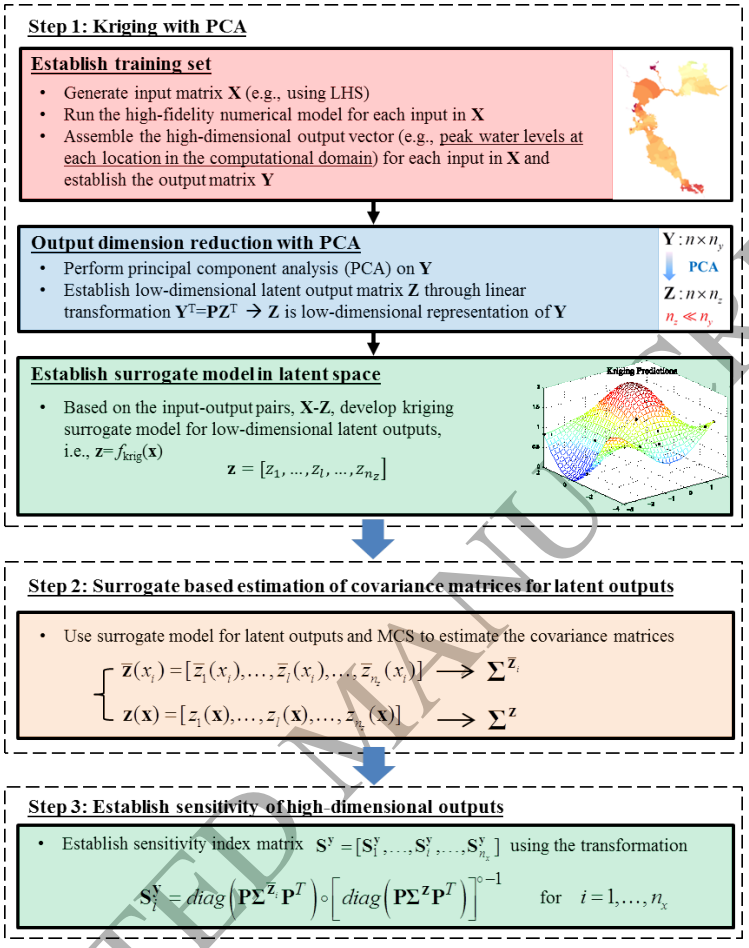


Figure 2: Flow chart for DRESSA for expensive models with high-dimensional outputs.

putational efficiency [16, 7, 8]. First, we establish a database with n total runs of the high-fidelity models. We will have the output vector $\{\mathbf{y}^h = \mathbf{y}(\mathbf{x}^h); h = 1, \ldots, n\}$ for each input $\{\mathbf{x}^h; h = 1, \ldots, n\}$. These inputs can be generated by Latin Hypercube Sampling (LHS). In the context of building surrogate models, LHS is commonly used to generate random samples of input parameter values that uniformly span the input parameter space so that the established surrogate model will have overall good performance over the entire input space. We will denote by $\mathbf{X} = [\mathbf{x}^1, \ldots, \mathbf{x}^n]^T \in R^{n \times n_x}$ and $\mathbf{Y} = [\mathbf{y}^1, \ldots, \mathbf{y}^n]^T \in R^{n \times n_y}$ the

corresponding input and output matrices, respectively. These observations are frequently referenced as *training set* or *support points*. For the surrogate model, we adopt the kriging surrogate modeling, known as the Best Linear Unbiased Predictor (BLUP), and for prediction it is based on only matrix manipulations. A generic illustration of kriging surrogate model response surface for a problem with two-dimensional inputs is shown in Figure 2.

For models with high-dimensional outputs, it is inefficient to train a surrogate model for each individual output. To address this, we adopt the approach in *Jia and Taflanidis* [7] where PCA, as a data-driven dimension reduction technique, is used to establish a low-dimensional latent output representation of the high-dimensional outputs. The data-driven nature of PCA means that it does not rely on the physical (e.g., spatial or temporal) correlation between different outputs, rather it exploits the correlation within the observation matrix $\mathbf{Y}$. PCA is established by considering the eigenvalue problem for the covariance matrix $\mathbf{Y}^T\mathbf{Y}$, and only the latent outputs and associated eigenvectors corresponding to the n_z largest eigenvalues are retained [9]. This leads to a $n \times n_z$ latent output matrix $\mathbf{Z}$ with the transformation between $\mathbf{Z}$ and $\mathbf{Y}$ defined as $\mathbf{Z}^T = \mathbf{P}^{-1}\mathbf{Y}^T$ where $\mathbf{P}$ is the $n_y \times n_z$ projection matrix containing the eigenvectors corresponding to the n_z largest eigenvalues. Value of n_z can be chosen so that latent outputs account for at least r_o [say 99%] of the total variance of the data [17]. Then we have $n_z \leq \min(n, n_y)$, with n_z being usually a small fraction of $\min(n, n_y)$. For $n << n_y$ (which is typically the case for high-dimensional outputs), we have $n_z << n_y$, leading to a significant reduction in the output dimension. Let $\mathbf{z} = [z_1, \ldots, z_{n_z}]$ denote the vector of latent outputs. Due to the low dimensionality of the latent outputs $\mathbf{z}$, a kriging surrogate model is built with respect to $\mathbf{z}$, based on the observation matrix $\mathbf{Z}$ for the input matrix $\mathbf{X}$.

More specifically, based on $\mathbf{X}$ and corresponding latent output matrix $\mathbf{Z}$, kriging establishes an approximation/prediction to $\mathbf{z}(\mathbf{x})$, denoted $\hat{\mathbf{z}}(\mathbf{x})$, for any new input $\mathbf{x}$ through $\hat{\mathbf{z}}(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T\boldsymbol{\alpha}^* + \mathbf{r}(\mathbf{x})^T\boldsymbol{\beta}^*$, where $\boldsymbol{\alpha}^* = (\mathbf{F}^T\mathbf{R}^{-1}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{R}^{-1}\mathbf{Z}$

and $\boldsymbol{\beta}^* = \mathbf{R}^{-1}(\mathbf{Z} - \mathbf{F}\boldsymbol{\alpha}^*)$ are the $n_p \times n_z$ dimensional and $n \times n_z$ dimensional, respectively, coefficient matrices. Here $\mathbf{f}(\mathbf{x})$ is the n_p-dimensional basis vector (e.g., linear or quadratic polynomials of $\mathbf{x}$), and $\mathbf{F} = [\mathbf{f}(\mathbf{x}^1)\dots\mathbf{f}(\mathbf{x}^n)]^T$ is the $n \times n_p$ basis matrix. $\mathbf{R}$ is the $n \times n$ correlation matrix with the jk^{th} element defined as $R(\mathbf{x}^j,\mathbf{x}^k)$. Here the generalized exponential correlation is used for $R(.)$, where the optimal selection of the parameters of the correlation function is established using the Maximum Likelihood Estimation (MLE) principle. More details about kriging surrogate model can be found in *Jia and Taflanidis* [7]. For prediction at any new input $\mathbf{x}$, kriging will be used in place of the high-fidelity model. When the output $\mathbf{y}$ is the interested quantity, the transformation $\mathbf{y}^T = \mathbf{P}\mathbf{z}^T$ can be used directly to establish prediction for $\mathbf{y}$, i.e., $\hat{\mathbf{y}}(\mathbf{x})^T = \mathbf{P}\hat{\mathbf{z}}(\mathbf{x})^T$. Since here we are interested in the sensitivity index (which involves calculation of the variance of $\mathbf{y}$), we need to establish the corresponding transformation for sensitivity index between latent and initial output spaces, which is discussed next.

2.3.2. Efficient Sensitivity Analysis for High-dimensional Outputs

To efficiently carry out sensitivity analysis for high-dimensional output, we derive the relationship between the sensitivity indexes of the high-dimensional original outputs with respect to the inputs and the relevant covariance matrices of the much lower-dimensional latent outputs. This is one of the novelties of the proposed method, compared to the kriging with PCA in *Jia and Taflanidis* [7] where the transformation is only established for the responses. More specifically, for sensitivity analysis of high-dimensional outputs, we first calculate the covariance matrices for the low-dimensional latent outputs using the established kriging model for the latent outputs. Then we use the derived relationship to directly calculate the sensitivity indexes for the high-dimensional original outputs. Next, some details about the transformation are provided.

Let $\bar{y}_k(x_i)$ represent $E_{\sim i}\left(y_k(\mathbf{x})|x_i\right)$ and similarly for the latent outputs

$\bar{z}_l(x_i)$ represent $E_{\sim i}\left(z_l(\mathbf{x})|x_i\right)$. Based on the transformation $\mathbf{y}(\mathbf{x})^T = \mathbf{P}\mathbf{z}(\mathbf{x})^T$, we can express the variances of $\bar{y}_k(x_i)$ and $y_k(\mathbf{x})$ in terms of the variances of the latent outputs $\bar{\mathbf{z}}_i = \bar{\mathbf{z}}(x_i) = [\bar{z}_1(x_i), \ldots, \bar{z}_l(x_i), \ldots, \bar{z}_{n_z}(x_i)]$ and $\mathbf{z} = \mathbf{z}(\mathbf{x}) = [z_1(\mathbf{x}), \ldots, z_l(\mathbf{x}), \ldots, z_{n_z}(\mathbf{x})]$, respectively. Let $\mathbf{\Sigma}^{\mathbf{z}}$ denote the $n_z \times n_z$ covariance matrix for the latent outputs $\mathbf{z}(\mathbf{x})$, and $\mathbf{\Sigma}^{\bar{\mathbf{z}}_i}$ denote the $n_z \times n_z$ covariance matrix for $\bar{\mathbf{z}}(x_i)$. The covariance matrices of $\bar{\mathbf{y}}_i = \bar{\mathbf{y}}(x_i) = [\bar{y}_1(x_i), \ldots, \bar{y}_k(x_i), \ldots, \bar{y}_{n_y}(x_i)]$ and $\mathbf{y} = \mathbf{y}(\mathbf{x}) = [y_1(\mathbf{x}), \ldots, y_k(\mathbf{x}), \ldots, y_{n_y}(\mathbf{x})]$, can be written as $\mathbf{\Sigma}^{\bar{\mathbf{y}}_i} = \mathbf{P}\mathbf{\Sigma}^{\bar{\mathbf{z}}_i}\mathbf{P}^T$ and $\mathbf{\Sigma}^{\mathbf{y}} = \mathbf{P}\mathbf{\Sigma}^{\mathbf{z}}\mathbf{P}^T$, respectively. Then we derive the following relationship between the sensitivity indexes for the original outputs and the covariance matrices in the latent space, $\mathbf{S}_i^{\mathbf{y}} = diag(\mathbf{P}\mathbf{\Sigma}^{\bar{\mathbf{z}}_i}\mathbf{P}^T) \circ \left[diag(\mathbf{P}\mathbf{\Sigma}^{\mathbf{z}}\mathbf{P}^T)\right]^{\circ -1}$, where $\mathbf{S}_i^{\mathbf{y}}$ corresponds to the sensitivity indexes of all the outputs with respect to the i^{th} input x_i (i.e., the i^{th} column of $\mathbf{S}^{\mathbf{y}}$) and $diag(.)$ denotes the diagonal operator and establishes a vector consisting of the diagonal elements of a matrix. The notation $A \circ B$ means the Hadamard product of matrix A and B, i.e., elementwise product, and $B^{\circ -1}$ means the Hadamard inverse of matrix B, i.e., elementwise inverse. We can see that to calculate $\mathbf{S}_i^{\mathbf{y}}$ (for $i = 1, \ldots, n_x$), we can first calculate the covariance matrices $\mathbf{\Sigma}^{\bar{\mathbf{z}}_i}$ (for $i = 1, \ldots, n_x$) and $\mathbf{\Sigma}^{\mathbf{z}}$, and then use the above transformation to establish the sensitivity index for the original high-dimensional outputs. These covariance matrices (i.e., $\mathbf{\Sigma}^{\bar{\mathbf{z}}_i}$ and $\mathbf{\Sigma}^{\mathbf{z}}$) can be estimated using the established kriging surrogate model $\hat{\mathbf{z}}(\mathbf{x})$ for the latent outputs $\mathbf{z}(\mathbf{x})$ and MCS. This way, the size of matrices to be kept in memory for sensitivity analysis is reduced to the order of $N \times n_z \times n_x$, corresponding to a reduction of n_y/n_z compared to $N \times n_y \times n_x$. For the example discussed later, $n_y/n_z \approx 100000/20 = 5000$, which means 5000 times reduction in memory requirements. The reduction in memory and matrix size will also lead to significant improvement in computational efficiency. The sensitivity indexes for higher order interactions can be derived similarly. Overall, the proposed approach allows efficient investigation of the sensitivity of expensive models with high-dimensional outputs. In the current case, we use the proposed approach to efficiently generate sensitivity maps.

3. Example and Implementation Details

Here we consider building containments at the county level, and 8 counties are considered. The input $\mathbf{x}$ has dimension of $n_x = 8$ where x_i for $i = 1, \ldots, n_x$ corresponds to Marin County, Sonoma County, Napa County, Solano County, Contra Costa County, Alameda County, Santa Clara County, and San Mateo County, respectively. Under the projected SLR, the high-fidelity model discussed in Section 2.1 is adopted to simulate the response of SF Bay region for any selected input $\mathbf{x}$. When $x_i \geq 0.5$, the seawall at the corresponding county will be constructed, which will be modeled in the high-fidelity model by inserting a flow barrier at the seawall location. For the projected SLR, we select the case of 1.5m. All high-fidelity runs will be under 1.5m of SLR. To establish the database for building a surrogate model, we generate n samples for the input and evaluate the high-fidelity model for each of the n inputs, giving the corresponding output matrix $\mathbf{Y}$. For some of the locations, they remain dry for some or all of the inputs, which creates problems for both establishing latent outputs using PCA and building a surrogate model due to the discontinuity in the water levels; theses dry nodes are not included in the outputs. This results to output with dimension of $n_y = 80,050$.

For selection of n, it is expected that larger values will lead to better accuracy of kriging. However, due to the high computational effort of the high-fidelity model, it is desirable to use small number for n as long as certain accuracy is reached. Here we use the relative mean absolute error ($RMAE$) to quantify the approximation errors of kriging model. The $RMAE$ for original output $\mathbf{y}$ averaged over all locations (i.e., $ARMAE = \sum_{k=1}^{n_y} RMAE_k/n_y$) is adopted, since we are ultimately interested in having small error for the original outputs. Another option is to calculate the error for the selected latent outputs. However, it is expected that the errors for the latent outputs and the errors for the original outputs will be consistent. To calculate $RMAE_k$ for the k^{th} output, we use leave-one-out cross validation. This cross-validation is performed as follows: first,

each of the observations from the database is sequentially removed, then the remaining support points are used to predict the output for this removed support point, and the error between the predicted and real outputs is evaluated. The cross-validation error $RMAE_k$ is obtained by averaging the errors established over all observations, i.e., $RMAE_k = \sum_{h=1}^{n} |y_k(\mathbf{x}^h) - \hat{y}_k(\mathbf{x}^h)| / \sum_{h=1}^{n} |y_k(\mathbf{x}^h)|$. When n is larger than 40, there is little variation in $ARMAE$, and the $ARMAE$ for $n = 40$ is around 0.77%, which is below the targeted accuracy level of 1% $ARMAE$, indicating good accuracy for the established surrogate model. Therefore, we use $n = 40$ for building the kriging model. For the number of latent outputs for PCA, we select $r_o = 99.9\%$ (which means the error due to PCA truncation is negligible), and the corresponding number of required latent outputs is $n_z = 20$. We have $n_z << n_y = 80,050$, which leads to significant improvement in computational efficiency and reduction in memory requirements.

Then the surrogate model built with $n = 40$ high-fidelity runs is used within MCS for calculation of variances and sensitivity indexes (corresponding to Steps 2 and 3 in the flowchart in Figure 2) for both first order main effects and second order interactions, which are discussed next.

4. Results and Discussions

4.1. Sensitivity Map for First Order Main Effects

The sensitivity maps for the first order main effect are shown in Figure 3. From Figure 3(f) and (h), it can be seen that the construction of containments at Alameda County or San Mateo County will have large impact on the variation of PWL in the entire bay. On the other hand, from Figure 3(a), (b), and (e), it can be seen that the construction of containments at Marin County, or Sonoma County, or Contra Costa County will have very small impact on the variation of PWL in the entire bay. These results of strong and weak effects are consistent with the local sensitivity study of *Wang et al.* [21]. In their study,

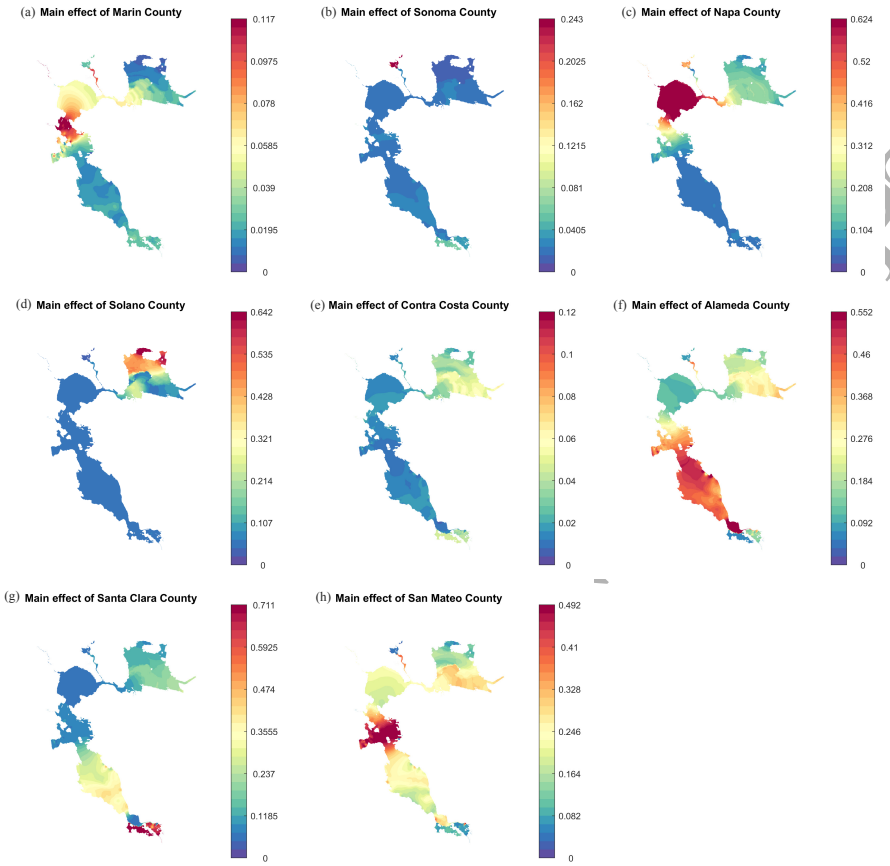


Figure 3: Variance-based sensitivity map with respect to x_i (corresponding to each county)

the shoreline is altered one by one and when an alteration is made all the other counties' shoreline remain unprotected. Similarly, they found that Alameda and San Mateo Counties have strong hydrodynamics impacts than the others. Particularly, we found strong similarity between our Figure 3 and their Figure 5.

As to Napa County, Figure 3(c) shows that the construction of containments at Napa County will have relatively large impact on the variation of PWL in the San Pablo Bay (SPB) region while having small impact on the PWL of the rest of the bay. Figure 3(d) shows that the construction of containments at Solano

County will have localized impacts on the PWL, that is, large impact on northern parts of Suisun Bay (SB) while having relatively small impact on the PWL of the rest of the bay. Figure 3(g) shows that the construction of containments at Santa Clara County will have localized impacts on the PWL in southern part of the South SF Bay (SSFB) but at the same time moderate level of impacts on the rest of the bay. The main effects of these counties are shown more regional in the present global sensitivity study than the local sensitivity of *Wang et al.* [21]. This indicates that the local sensitivity analysis underestimates the scale of the impacts.

Note that the upstream Napa River shows an outstanding sensitivity to the shoreline alteration. The reason is that the spot is located at the channel bifurcation/convergence that is formed by the configuration of the levees. So the water level increase in San Pablo Bay can propagate to this sensitive location through the converging channels. Consequently, this converging effect amplifies the water level change in San Pablo Bay and lead to the high sensitivity at this spot.

It is important to note that the purpose of the sensitivity maps (i.e., showing the sensitivity index values) is to investigate the relative importance of each input (i.e., building containment at corresponding county or not) on the variability of output (i.e., PWL) at different locations on the map to guide decision making regarding whether to build containment at specific counties and the potential impacts over the entire bay. On the other hand, if the actual (or absolute) variability/variance is the quantity of interest, the sensitivity map can be easily converted to a variance map by multiplying the total variance V_y at each location, e.g., $V_i = S_i V_y$. This V_y can be established using the transformation $\mathbf{\Sigma^y} = \mathbf{P}\mathbf{\Sigma^z}\mathbf{P}^T$ mentioned earlier and V_y corresponds to the diagonal elements of $\mathbf{\Sigma^y}$. Again $\mathbf{\Sigma^z}$ can be calculated using the established surrogate model. In such case, the variability V_i of the output (i.e., meters of change for the PWL) due to x_i can be directly read from the variance map.

4.2. Sensitivity Map for Second Order Interactions

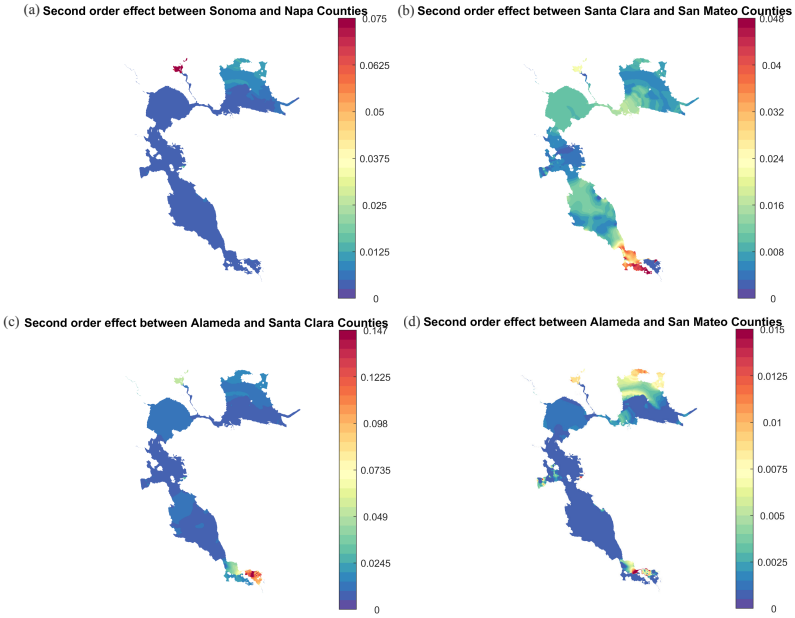


Figure 4: Sensitivity map for second order interactions

In interpreting the sensitivity results $S_{[ij]}$ for the second order interactions between two inputs x_i and x_j, it is important to differential such results from the joint first order main effects S_{ij} for the same two inputs, which based on the variance decomposition discussed in Section 2.2 corresponds to $S_{ij} = S_i + S_j + S_{[ij]}$, i.e., summation of first order main effects due to x_i and x_j and the second order interaction between x_i and x_j. In this study, under the considered SLR scenario, overall, it was found that the second order interactions are very weak, with sensitivity values close to zero. Figure 4 shows the second order interactions between several counties that have relatively large (but still with small sensitivity index values, in the order of 0.01) but very localized impacts on some regions, while the rest of the 24 second order sensitivity indexes for the entire region is close

to 0 and is not shown. Figure 4(b) shows that the second order interaction between Alameda and Santa Clara Counties will have relatively large (still small, e.g., in the order of 0.01) impact on the south and north tips. Overall, the second order interaction information is useful when considering the impact of construction of containments at two counties. These are the extra benefits of joint sensitivity of the variance-based global sensitivity analysis, which the local sensitivity analysis cannot provide.

5. Conclusions

We applied the Dimension REduced Surrogate based Sensitivity Analysis (DRESSA) method to generate sensitivity maps and investigate the impact of different containment strategies on PWL over the entire SF Bay under SLR. Regional impacts were strongest for actions by SSFB counties (Alameda and San Mateo) and for the county just north of the Golden Gate channel (Marin). In each case, this regional sensitivity results from changes to the basin-scale tidal dynamics: for the south bay counties, their shoreline interventions ensure amplification of the tides in South Bay which feeds back to the Northern part of the Bay; for Marin county, the intervention promotes north-to-south information transfer through that part of the Bay. The results of the main effects show that the global sensitivity analysis reveals more regional impacts than the local sensitivity analysis. Overall, the second order interactions are found to be weak, and neglecting them leads to errors of less than 10%. Although higher order interactions were negligible for these analyses of actions at the scale of counties, it is expected that if the scale of action were reduced, the primary effects would be weakened and second-order (or third-order) effects would become more important. The DRESSA method makes it computationally possible to disaggregate the shorelines into a larger number of segments and it can be easily extended to other regions. It is important to note that the DRESSA method is general and can be easily extended to other regions.

425

Acknowledgement

This work is supported by National Science Foundation grant (1541181). This support is gratefully acknowledged. Data is available through https://figshare.com/s/ded1fb1f1a8a3e686271.

Appendix: Nomenclature

$\mathbf{x}$	Design variables (inputs)
x_i	i^{th} design variable
n_x	Number of design variables
$\mathbf{y}$	Output vector
n_y	Dimension of output vector
V_y	Total variance of y due to uncertainty in $\mathbf{x}$
V_i	Expected reduction in variance of y due to fixing x_i
S_i	First order Sobol' index for x_i
$S_{[ij]}$	Sobol' index for second order interaction between x_i and x_j
$S^y_{k,i}$	First order Sobol' index of k^{th} output y_k w.r.t the i^{th} input x_i
$\mathbf{S^y}$	$n_y \times n_x$ sensitivity index matrix
$\mathbf{S}^{\mathbf{y}}_i$	Sensitivity indexes of all the outputs with respect to the i^{th} input x_i, i.e., the i^{th} column of $\mathbf{S^y}$
N	Number of samples for Monte Carlo Simulation based estimation of Sobol' indexes
$\mathbf{X}$	Input matrix of the training set for surrogate modeling
$\mathbf{Y}$	Output matrix of the training set for surrogate modeling
n	Number of samples in the training set
$\mathbf{Z}$	Latent output matrix of the training set for surrogate modeling
$\mathbf{P}$	Transformation matrix between original output and latent output
$\mathbf{z}$	Latent output vector
n_z	Number of latent outputs
$\hat{\mathbf{z}}(\mathbf{x})$	Kriging prediction of $\mathbf{z}(\mathbf{x})$ at $\mathbf{x}$
$\mathbf{f}(\mathbf{x})$	Basis vector for kriging surrogate model
$\mathbf{F}$	Basis matrix for kriging surrogate model
$\mathbf{R}$	$n \times n$ correlation matrix for kriging surrogate model
$\bar{z}_l(x_i)$	$\bar{z}_l(x_i) = E_{\sim i}\left(z_l(\mathbf{x})\|x_i\right)$
$\bar{\mathbf{z}}_i$	$\bar{\mathbf{z}}_i = \bar{\mathbf{z}}(x_i) = [\bar{z}_1(x_i), \ldots, \bar{z}_l(x_i), \ldots, \bar{z}_{n_z}(x_i)]$
$\Sigma^{\bar{\mathbf{z}}_i}$	$n_z \times n_z$ covariance matrix of $\bar{\mathbf{z}}_i$
$\Sigma^{\mathbf{z}}$	$n_z \times n_z$ covariance matrix of $\mathbf{z}$

References

References

[1] Bertin, X., K. Li, A. Roland, Y. J. Zhang, J. F. Breilh, and E. Chaumillon (2014), A modeling-based analysis of the flooding associated with xynthia, central bay of biscay, *Coastal Engineering*, *94*, 80–89.

[2] Deltares (2016), D-flow flexible mesh, Retrieved from https://www.deltares.nl/en/software/delft3d-flexible-mesh-suite/.

[3] Doehring, C., J. Beagle, J. Lowe, R. M. Grossinger, M. Salomon, P. Kauhanen, S. Nakata, R. A. Askevold, and S. N. Bezalel (2016), San Francisco Bay Shore Inventory: Mapping for Sea Level Rise Planning, *Tech. rep.*, San Francisco Estuary Institute, Richmond, CA.

[4] Elias, E. P., and J. E. Hansen (2013), Understanding processes controlling sediment transports at the mouth of a highly energetic inlet system (San Francisco Bay, CA), *Marine Geology*, *345*, 207–220.

[5] Erikson, L. H., S. A. Wright, E. Elias, D. M. Hanes, D. H. Schoellhamer, and J. Largier (2013), The use of modeling and suspended sediment concentration measurements for quantifying net suspended sediment transport through a large tidally dominated inlet, *Marine Geology*, *345*, 96–112.

[6] Holleman, R. C., and M. T. Stacey (2014), Coupling of Sea Level Rise, Tidal Amplification, and Inundation, *Journal of Physical Oceanography*, *44*(5), 1439–1455.

[7] Jia, G., and A. A. Taflanidis (2013), Kriging metamodeling for approximation of high-dimensional wave and surge responses in real-time storm/hurricane risk assessment, *Computer Methods in Applied Mechanics and Engineering*, *261-262*, 24–38.

[8] Jia, G., A. A. Taflanidis, N. C. Nadal-Caraballo, J. A. Melby, A. B. Kennedy, and J. M. Smith (2016), Surrogate modeling for peak or time-dependent storm

surge prediction over an extended coastal region using an existing database of synthetic storms, *Natural Hazards*, *81*(2), 909–938.

[9] Jolliffe, I. T. (2002), *Principal Component Analysis*, 488 pp.

[10] Lee, S. B., M. Li, and F. Zhang (2017), Impact of sea level rise on tidal range in chesapeake and delaware bays, *Journal of Geophysical Research: Oceans*, *122*(5), 3917–3938.

[11] Martyr-Koller, R. C., H. W. Kernkamp, A. van Dam, M. van der Wegen, L. V. Lucas, N. Knowles, B. Jaffe, and T. A. Fregoso (2017), Application of an unstructured 3D finite volume numerical model to flows and salinity dynamics in the San Francisco Bay-Delta, *Estuarine, Coastal and Shelf Science*, *192*, 86–107.

[12] NOAA (2016), NOAA tides and currents, Retrieved from https://tidesandcurrents.noaa.gov/.

[13] Passeri, D. L., S. C. Hagen, N. G. Plant, M. V. Bilskie, S. C. Medeiros, and K. Alizad (2016), Tidal hydrodynamics under future sea level rise and coastal morphology in the northern gulf of mexico, *Earth's Future*, *4*(5), 159–176.

[14] Pelling, H. E., and J. M. Green (2014), Impact of flood defences and sea-level rise on the european shelf tidal regime, *Continental Shelf Research*, *85*, 96–105.

[15] Sobol', I. (2001), Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Mathematics and Computers in Simulation*, *55*(1-3), 271–280.

[16] Taflanidis, A. A., G. Jia, A. B. Kennedy, and J. M. Smith (2013), Implementation/optimization of moving least squares response surfaces for approximation of hurricane/storm surge and wave responses, *Natural Hazards*, *66*(2), 955–983.

[17] Tipping, M., and C. Bishop (1999), Probabilistic principal component analysis, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *61*(3), 611–622.

[18] USGS (2016), USGS current water data for the nation, Retrieved from https://waterdata.usgs.gov/nwis/rt.

[19] Wang, R.-Q., E. Adams, A. W. K. Law, and A. Lai (2015), Scaling particle cloud dynamics: from lab to field, *Journal of Hydraulic Engineering*, *141*, 06015006.

[20] Wang, R.-Q., L. M. Herdman, L. Erikson, P. Barnard, M. Hummel, and M. T. Stacey (2017), Interactions of Estuarine Shoreline Infrastructure With Multiscale Sea Level Variability, *Journal of Geophysical Research: Oceans*, *122*, 9962–9979.

[21] Wang, R.-Q., M. T. Stacey, L. M. M. Herdman, P. L. Barnard, and L. Erikson (2018), The influence of sea level rise on the regional interdependence of coastal infrastructure, *Earth's Future*, *6*, https://doi.org/10.1002/2017EF000742